

KLASIFIKASI MONKEYPOX MENGGUNAKAN EKSTRAKSI FITUR GLCM DAN ALGORITMA RANDOM FOREST

William Wijaya^{*1)}, Muhammad Rizky Pribadi², Eka Puji Widiyanto³

^{1,2}Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

³Program Studi Teknik Elektro, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

¹willianswijaya@mhs.mdp.ac.id, ²rizky@mdp.ac.id, ³ekapujiw2002@mdp.ac.id

Kata kunci:

gray level co-occurrence matrix; monkeypox; random forest

Abstract: Monkeypox is a disease that was first discovered in 1958 in Denmark when there were two smallpox like cases in a colony of monkeys kept for research, hence the name 'Monkeypox'. GLCM is a matrix that represents the occurrence of two-pixel pair frequencies with gray intensity, distance and angle and Random Forest is a classification method which is a combination of every good tree then combined into one model. This study uses the public dataset monkeypox and non-monkeypox and has a total of 3,192 images divided into test data and training data with ratios of 60:40, 70:30, 80:20, and 90:10 and each image is resized to 224x224 pixels. The results of the GLCM texture features were tested using the Random Forest method using the parameters $n_estimator=100, 500, \text{ and } 1000$. From the four tests that have been carried out, it can be concluded that the best $n_estimator$ is 100 with a ratio proportion dataset of 90:10, which produces an accuracy value of 77%, precision 77%, recall 77%, and $f1\text{-score}$ 76.5%.

Abstrak: Monkeypox adalah penyakit yang pertama kali ditemukan pada tahun 1958 di Denmark ketika ada dua kasus seperti cacar pada koloni kera yang dipelihara untuk penelitian, sehingga cacar ini dinamakan 'Monkeypox'. GLCM adalah suatu matriks yang merepresentasikan frekuensi munculnya pasangan dua piksel dengan intensitas keabuan, jarak dan sudut dan Random Forest adalah salah satu metode klasifikasi yang merupakan kombinasi dari setiap pohon yang baik kemudian dikombinasikan ke dalam satu model. Penelitian ini menggunakan public dataset monkeypox dan non-monkeypox dan memiliki gambar berjumlah 3.192 dibagi menjadi data uji dan data latih dengan rasio 60:40, 70:30, 80:20, dan 90:10. Hasil dari fitur tekstur GLCM dilakukan pengujian dengan metode Random Forest menggunakan parameter $n_estimator = 100, 500, \text{ dan } 1000$. Dari empat pengujian yang sudah dilakukan, disimpulkan bahwa $n_estimator$ terbaik adalah 100 dengan proporsi rasio dataset 90:10, yang menghasilkan nilai accuracy sebesar 77%, precision 77%, recall 77%, dan $f1\text{-score}$ 76,5%.

Wijaya dkk. (2023). Klasifikasi Monkeypox Menggunakan Ekstraksi Fitur GLCM dan Algoritma Random Forest. MDP Student Conference 2023.

PENDAHULUAN

Monkeypox (cacar monyet) adalah penyakit yang pertama kali ditemukan pada tahun 1958 di Denmark ketika ada dua kasus seperti cacar pada koloni kera yang dipelihara untuk penelitian, sehingga cacar ini dinamakan 'Monkeypox'. Penularan Monkeypox pada manusia terjadi melalui kontak langsung dengan orang ataupun hewan yang terinfeksi, selain itu bisa juga melalui benda yang terkontaminasi oleh virus tersebut [1].

Cacar monyet dan cacar air memiliki gejala yang mirip, yaitu ruam pada kulit mulai lesi datar berwarna merah, benjolan kecil kemerahan seperti jerawat, lenting berisi cairan yang berubah warna menjadi keruh, hingga mengering dan membentuk keropeng [2].

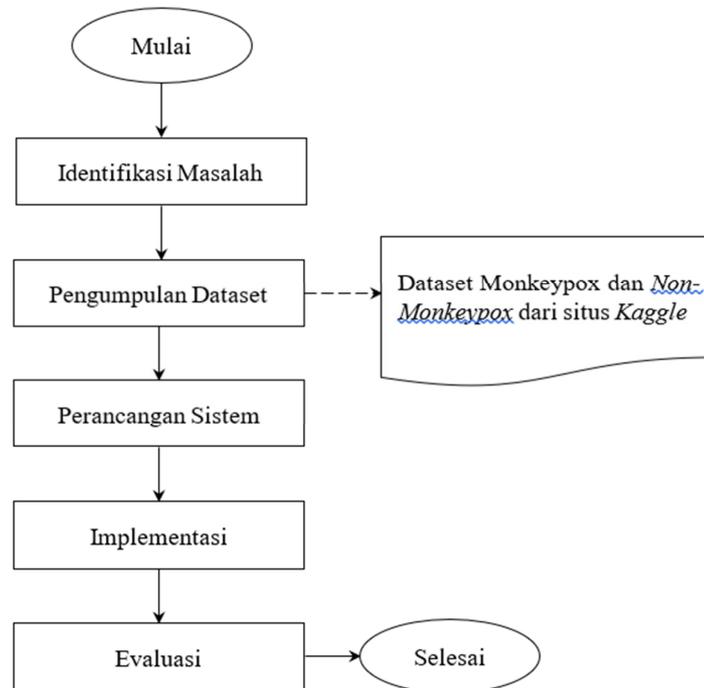
Sekarang ini banyak teknologi dikembangkan untuk mempermudah pekerjaan manusia, salah satunya adalah pengolahan citra digital. Pengolahan citra digital adalah disiplin ilmu yang mempelajari tentang teknik-teknik mengolah citra. Citra yang dimaksud disini adalah gambar diam (foto) maupun gambar bergerak (yang berasal dari webcam). Sedangkan digital mempunyai maksud bahwa pengolahan citra/gambar dapat dilakukan secara digital menggunakan computer [3].

Fitur tekstur adalah salah satu fitur pengolahan citra digital yang digunakan untuk melakukan ekstraksi sebuah objek, salah satunya adalah fitur *GLCM (Gray Level Co-occurrence Matrix)*. *GLCM* adalah suatu matriks yang merepresentasikan frekuensi munculnya pasangan dua piksel dengan intensitas keabuan, jarak dan sudut [4]. Ekstraksi ciri tekstur dengan *GLCM* 4 sudut, digunakan matriks kookurensi, matriks yang merepresentasikan hubungan ketetanggaan antar pixel pada citra dari berbagai arah orientasi dan jarak spasial [5]. Beberapa penelitian yang menggunakan fitur *GLCM* sudah diimplementasikan untuk klasifikasi citra *Magnetic Resonance Imaging (MRI)* otak dalam mengidentifikasi tumor, mammogram, dan objek lainnya. Penerapan fitur *GLCM* untuk klasifikasi citra *MRI* otak dalam mengidentifikasi tumor menghasilkan akurasi sebesar 87,64% [6]. Kemudian identifikasi virus *COVID-19* dari *chest x-ray* menggunakan fitur *GLCM* juga menghasilkan akurasi sebesar 99,17% [7].

Random Forest adalah salah satu metode klasifikasi yang merupakan kombinasi dari setiap pohon yang baik kemudian dikombinasikan ke dalam satu model. *Random forest* memiliki keunggulan yang dapat meningkatkan akurasi apabila terdapat data yang hilang serta untuk *resisting outliers*, dan juga efisien untuk penyimpanan data. Selain itu, *random forest* mempunyai proses seleksi fitur dimana proses ini mampu mengambil fitur terbaik untuk meningkatkan performa pada model klasifikasi. Adanya fitur seleksi membuat *random forest* mampu bekerja pada data yang besar dengan parameter yang kompleks secara efektif [8]. Penelitian klasifikasi kanker kulit menggunakan *Random Forest* menghasilkan akurasi sebesar 97,3% [9], dan penelitian mendiagnosis penyakit kanker payudara menghasilkan akurasi sebesar 99,51% [10].

Berdasarkan penjelasan penelitian sebelumnya, ekstraksi fitur *GLCM* memiliki tingkat akurasi yang baik untuk mengenali ciri tekstur pada objek, serta metode *Random Forest* juga memiliki tingkat akurasi yang baik dalam klasifikasi sebuah objek. Oleh karena itu akan dilakukan penelitian klasifikasi *Monkeypox* dan *Non-Monkeypox* menggunakan fitur ekstraksi *GLCM* dan algoritma *Random Forest*.

METODE



Gambar 1. Metode Penelitian

Klasifikasi Masalah

Tahap ini dimulai dengan mencari informasi dari jurnal, buku, artikel dan teori yang terkait dengan *monkeypox*. Selanjutnya mengumpulkan literatur sebagai referensi untuk memberikan kontribusi penelitian yang baru.

Pengumpulan Dataset

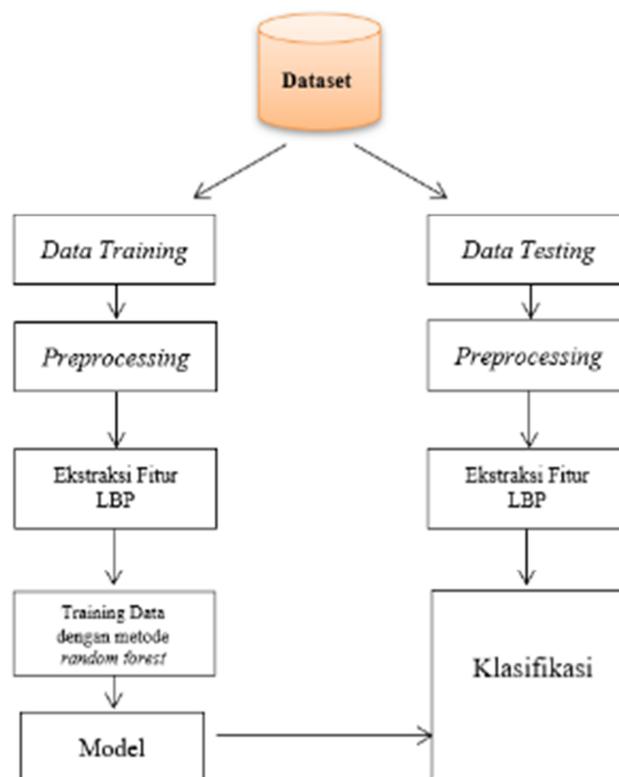
Dari dataset yang tersedia terdapat 2 kelas yakni citra monkeypox dan non-monkeypox (chickenpox dan measles). Dataset berjumlah 3.192 gambar yang akan dibagi dengan proporsi rasio 60:40, 70:30, 80:20, dan 90:10. Setiap gambar memiliki dimensi berukuran 224 x 224 pixels. Pembagian dataset dapat dilihat pada Tabel 1.

Tabel 1. Dataset *Monkeypox* dan *Non-Monkeypox*

No	Kelas	Citra	Jumlah Citra
1	<i>Monkeypox</i>		1428
2	<i>Non Monkeypox</i>		1764

Perancangan Sistem

Setelah mengumpulkan dataset, maka dilakukan Preprocessing. Pre-processing merupakan serangkaian proses untuk mempersiapkan citra sebelum proses segmentasi. Proses ini meliputi proses konversi citra RGB menjadi citra grayscale. Setelah semua citra menjadi grayscale, selanjutnya adalah membagi data dengan proporsi 90 (data training) : 10 (data test), 80:20, 70:30, dan 60:40. Langkah selanjutnya adalah melakukan proses ekstraksi fitur dengan pengambilan ciri dari citra *monkeypox* dan *non-monkeypox* yang menggambarkan karakteristik dari *monkeypox* dan *non-monkeypox*. Fitur yang di ambil dari citra *monkeypox* dan *non-monkeypox* adalah fitur tekstur. Pada proses pelatihan (training), citra *monkeypox* dan *non-monkeypox* akan dilakukan tahap ekstraksi fitur tekstur *GLCM*. Hasil dari citra akan digunakan untuk proses traning yang menggunakan metode *Random Forest*. Setelah proses training selesai, maka akan didapatkan data model yang akan digunakan untuk pengujian data. Pada proses pengujian (testing), citra *monkeypox* dan *non-monkeypox* dilakukan tahap ekstraksi fitur tekstur *GLCM*. Hasil dari citra akan digunakan untuk proses pengujian. Setelah pengujian selesai, maka akan didapatkan data yang mendekati atau sama dengan model yang sudah dibuat. Pengujian untuk model yang telah dibuat dengan menggunakan data uji yang telah melalui proses *thresholding*. *Dataset* yang digunakan telah melalui proses *resize* dan *grayscale* citra. Gambar 2 menunjukkan gambar perancangan sistem. Proses diatas diulangi hingga setiap rasio mendapatkan hasil pengujian yang selanjutnya digunakan untuk menghitung nilai *precision*, *recall*, *accuracy* dan *f-1 score*. Skema perancangan yang sudah dijelaskan dapat dilihat pada Gambar 2. Skema Perancangan Sistem.



Gambar 2. Skema Rancangan Sistem

Implementasi

Tahap ini mengimplementasikan perancangan sistem yang sudah dibuat menggunakan *Google Collab*.

Evaluasi

Pada tahap ini dilakukan evaluasi dari tahapan sebelumnya yaitu pengujian. Evaluasi yang menggunakan metode *Confusion Matrix* yang terdiri dari *Precision*, *Recall*, *f1-score* dan *Accuracy*. Untuk menghitung persentase *precision*, *recall*, dan *accuracy* dengan jumlah keluaran kelas yang lebih dari dua digunakan Persamaan (1), (2), (3), dan (4).

$$Precision = TP / TP + FP \times 100\% \quad (1)$$

$$Recall = TP / TP + FN \times 100\% \quad (2)$$

$$Accuracy = TP + TN / TP + FN + TN + FP \times 100\% \quad (3)$$

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

Dalam rumus di atas, TP adalah jumlah data positif citra yang terklasifikasi dengan benar. TN adalah jumlah data negatif citra yang terklasifikasi dengan benar. FN adalah jumlah data negatif citra namun terklasifikasi salah. FP adalah jumlah data positif citra namun terklasifikasi salah.

HASIL DAN PEMBAHASAN

Hasil dari *confusion matrix* selanjutnya dihitung untuk mendapatkan tingkat keberhasilan dari metode yang sudah ditentukan, dimana tingkat keberhasilan tersebut menampilkan nilai dari *Precision*, *Recall*, *Accuracy*, dan *F1 Score*.

Tabel 2. Hasil Pengujian dengan Proporsi 60:40

Kelas	n_estimator	Nilai			
		Precision	Recall	Accuracy	F1 Score
MONKEYPOX	100	76%	66%	74%	0,71
	500	75%	68%	75%	0,71
	1000	75%	68%	75%	0,71
NON MONKEYPOX	100	74%	82%	74%	0,77
	500	75%	81%	75%	0,77
	1000	74%	81%	75%	0,77

Berdasarkan Tabel 2, diketahui bahwa pada pengujian kelas *monkeypox* dan *non-monkeypox* dengan proporsi dataset 60:40 mendapatkan *n_estimator* terbaik 500 dan 1000 yang dilihat dari nilai *F1 Score* sebesar 71% untuk kelas *monkeypox* dan 77% untuk kelas *non-monkeypox* dengan nilai *Accuracy* tertinggi 75%.

Tabel 3. Hasil Pengujian dengan Proporsi 70:30

Kelas	n_estimator	Nilai			
		Precision	Recall	Accuracy	F1 Score
MONKEYPOX	100	78%	67%	76%	0,72
	500	78%	68%	77%	0,73
	1000	78%	66%	76%	0,72
NON MONKEYPOX	100	75%	84%	76%	0,79
	500	76%	84%	77%	0,80
	1000	75%	84%	76%	0,79

Berdasarkan Tabel 3, diketahui bahwa pada pengujian kelas *monkeypox* dan *non-monkeypox* dengan proporsi dataset 70:30 menghasilkan akurasi sebesar 77% pada $n_estimator$ 500. Namun kelas *monkeypox* mendapatkan *F1-Score* tertinggi sebesar 73%, sedangkan dengan kelas *non-monkeypox* mendapatkan *F1-Score* sebesar 80%.

Tabel 4. Hasil Pengujian Dengan Proporsi 80:20

Kelas	n_estimator	Nilai			
		Precision	Recall	Accuracy	F1 Score
MONKEYPOX	100	78%	64%	74%	0,70
	500	78%	64%	74%	0,71
	1000	78%	63%	74%	0,70
NON MONKEYPOX	100	72%	83%	74%	0,77
	500	72%	84%	74%	0,78
	1000	72%	84%	74%	0,77

Berdasarkan Tabel 4, diketahui bahwa pada pengujian kelas *monkeypox* dan *non-monkeypox* dengan proporsi dataset 80:20 mendapatkan akurasi yang sama yaitu sebesar 74%. Namun dilihat dari *f1-score*, $n_estimator$ 1000 menghasilkan nilai rata-rata *f1-score* lebih besar yaitu 74,5%.

Tabel 5. Hasil Pengujian dengan Proporsi 90:10

Kelas	n_estimator	Nilai			
		Precision	Recall	Accuracy	F1 Score
MONKEYPOX	100	82%	69%	77%	0,75
	500	81%	67%	75%	0,73
	1000	81%	67%	75%	0,73
NON MONKEYPOX	100	72%	85%	77%	0,78
	500	71%	84%	75%	0,77
	1000	71%	84%	75%	0,77

Berdasarkan Tabel 5, diketahui bahwa pada pengujian kelas *monkeypox* dan *non-monkeypox* dengan proporsi dataset 90:10 mendapatkan akurasi terbesar 77% pada $n_estimator = 100$ dengan nilai *f1-score* sebesar 75% untuk kelas *monkeypox* dan 78% untuk kelas *non-monkeypox*.

KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan pada penelitian ini, didapatkan beberapa kesimpulan akhir diantaranya adalah fitur ekstraksi tekstur *GLCM* dan metode *random forest* dapat digunakan untuk klasifikasi *monkeypox* dan *non-monkeypox*. Dari empat pengujian yang sudah dilakukan, disimpulkan bahwa $n_estimator$ terbaik adalah 100 dengan proporsi rasio dataset 90:10, yang menghasilkan nilai *accuracy* sebesar 77%, nilai rata-rata *precision* 77%, nilai rata-rata *recall* 77%, dan nilai rata-rata *f1-score* 76,5%, dimana kelas *non-monkeypox* berhasil terprediksi sebagai kelas *non-monkeypox* sebanyak 134 dan 24 terprediksi sebagai kelas *monkeypox*. Sedangkan untuk kelas *monkeypox* hanya terprediksi 111 sebagai kelas *monkeypox*, dan 51 terprediksi sebagai kelas *non-monkeypox*.

DAFTAR PUSTAKA

- [1] R. Sinto, Shakinah Sharifah, M. Epid, Ibrahim, and L. Mendra, “Pedoman Pencegahan dan Pengendalian Penyakit Monkeypox,” 2022. [Online]. Available: <http://infeksiemerging.kemkes.go.id>
- [2] J. Ng, “Perbedaan Cacar Monyet dan Cacar Air,” 2021. <https://aido.id/health-articles/perbedaan-cacar-monyet-dan-cacar-air/detail> (accessed Dec. 13, 2022).
- [3] S. T. E. Mulyanto, V. Suhartono, O. D. Nurhayati, and Wijinarto, *Teori Pengolahan Citra Digital*. Semarang: Andi Yogyakarta dan UDINUS Semarang, 2009.
- [4] R. Widodo *et al.*, “Pemanfaatan Ciri Gray Level Co-Occurrence Matrix (GLCM) Citra Buah Jeruk Keprok (*Citrus reticulata* Blanco) Untuk Klasifikasi Mutu,” 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] N. Abdul Haris, “Kombinasi Ciri Bentuk dan Ciri Tekstur Untuk Identifikasi Penyakit pada Tanaman Padi,” *Kombinasi Ciri Bentuk dan Ciri Tekstur Untuk Identifikasi Penyakit Pada Tanaman Padi*, Vol. 7, No. 2, pp. 237–250, Aug. 2020.
- [6] Z. Fatni, “Klasifikasi Citra Magnetic Resonance Imaging (MRI) Otak Dalam Mengidentifikasi Tumor Menggunakan Algoritma Random Forest,” 2021.
- [7] S. D. Thepade, S. v. Bang, P. R. Chaudhari, and M. R. Dindorkar, “Covid19 Identification from Chest X-ray Images Using Machine Learning Classifiers with GLCM Features,” *Electronic Letters on Computer Vision and Image Analysis*, Vol. 19, No. 3, pp. 85–97, 2020, doi: 10.5565/REV/ELCVIA.1277.
- [8] S. Devella, Y. Yohannes, and F. N. Rahmawati, “Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.
- [9] S. Nandhini, M. A. Sofiyan, S. Kumar, and A. Afridi, “International Journal of Management and Humanities (IJMH) Skin Cancer Classification Using Random Forest,” *International Journal of Management and Humanities (IJMH)*, Vol. 4, No. 3, pp. 39–42, 2019, doi: 10.35940/ijmh.C0434.114319.
- [10] V. Angkasa and J. J. Pangaribuan, “Komparasi Tingkat Akurasi Random Forest dan KNN Untuk Mendiagnosis Penyakit Kanker Payudara,” 2022.